Providing an extended use of an otological-specific outcome instrument to derive cost-effectiveness estimates of treatment

Frew, Emma; Harrison, Mark; Rossello Roig, Melcior; Martin, Tom

DOI:
10.1111/coa.12424

Document Version
Peer reviewed version

Citation for published version (Harvard):
Frew, E, Harrison, M, Rossello Roig, M & Martin, T 2015, 'Providing an extended use of an otological-specific outcome instrument to derive cost-effectiveness estimates of treatment', *Clinical Otolaryngology*, vol. 40, no. 6, pp. 593-599. https://doi.org/10.1111/coa.12424

Full title: Providing an extended use of an otological-specific outcome instrument to derive cost-effectiveness estimates of treatment

Short title: Generating QALYs from an otological-specific outcome instrument

Frew, E.J. [1], Harrison, M.[2,3], Rossello Roig, M.[1], Martin, T.P.C.[4]

[1] Health Economics Unit, School of Health and Population Sciences, Public Health Building, University of Birmingham, Birmingham, B15 2TT, United Kingdom.
[2] Faculty of Pharmaceutical Sciences, University of British Columbia, 2405 Wesbrook Mall, Vancouver, BC V6T 1Z3, Canada
[3] Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, 588-1081 Burrard St, Vancouver, BC V6Z 1Y6, Canada
[4] ENT Department, Worcester Royal Hospital, Charles Hastings Way, Worcester, WR5 1DD, United Kingdom.

*Correspondence to: E Frew: e.frew@bham.ac.uk. Tel: +44 (0)121 414 3199, Fax: +44 (0) 121 414 8969.

ABSTRACT

Objectives: Although Quality-Adjusted Life Years (QALYs) are increasingly being used by decision makers to make comparisons of cost-effectiveness, there are no otological-specific outcome measures that fit within this QALY framework. This study had two main objectives. The first was to provide a means to derive QALYs from a condition-specific otological instrument (COQOL) and the second, was to assess the convergent validity, or degree of correlation, between the COQOL and SF-6D, an established QALY instrument.

Design: Longitudinal cohort study designed to assess the convergent validity between SF-6D and COQOL, and to generate a mapping function to enable SF-6D values to be predicted from the COQOL responses.

Setting: Cambridge University Hospital, UK

Participants: 207 patients attending a routine outpatient general otology clinic.

Main outcome measures: SF-6D and the COQOL instrument completed at baseline, and again 3 months later.

Results: Convergent validity was demonstrated with mean SF-6D values decreasing linearly with increasing severity on the COQOL instrument. Overall, the correlation between the COQOL scores and the SF-6D values was moderate and statistically significant ($r=.490$, $p=<0.001$). A simple mapping model based on an Ordinary Least Squares regression function predicted SF-6D values from the COQOL data with a reasonable degree of accuracy. Further validation using the follow up 3-month data confirmed the prediction power of this mapping model.

Conclusions: This study provides a method for estimating QALYs from condition-specific COQOL data and provides the opportunity for the cost-effectiveness of otological treatment to be measured and placed within the national QALY framework.

Health care resources are scarce and difficult decisions are required to allocate expenditure in a manner that is both efficient and equitable. To enable the assessment of the merits of alternative treatment options across different clinical areas, a common unit of health outcome is required, and in the UK, decision making bodies such as the National Institute of Health and Care Excellence (NICE) recommend the use of the Quality-Adjusted Life Year (QALY)[1].

QALYs are constructed by combining information on length of life with an index value representing a person's health-related quality of life (HRQOL) on a cardinal scale with death (0) and full health (1) denoting either end of the scale. Thus a person's 'QALY-expectancy' is the life expectancy 'adjusted' for the quality of life experienced. For example, if a person has an index value for HRQOL equal to 0.5, and 10 years left to live, then the QALY-expectancy is equal to 5 QALYs (10*0.5). If treatment improves the index value for HRQOL to 0.6 or extends life by 2 years, the QALY-expectancy would increase to 6 (0.6*10, or 0.5*12), and the QALY gain attributable to treatment under both scenarios would be 1 QALY. Overtime, the use of QALYs as a decision making tool has increased and is now used to inform resource allocation decisions worldwide[2].

Generic preference-based instruments have been designed to generate index values for the estimation of QALYs, such as the EQ-5D[3] or the SF-6D[4]. Generic instruments are applied across a range of conditions, allowing decision makers to compare the return in health outcomes for investments in different clinical areas on a common scale. These instruments are completed by patients and describe their current health state on the domains covered by the instrument: this combination of domain scores is known as their health profile. Each health profile in a preference-based measure has an index value attributed to it, which represents the preferences of a sample of respondents (usually a general population sample) for different health profiles. The final index value reflects a weighting of the domains within the instruments, for example 'pain' might be weighted more than 'self-care', and it is this preference-based weighting that makes HRQOL index values distinct from general HRQOL scores.

It is generally regarded as good practice for clinical studies to use condition-specific HRQOL instruments as these have been developed to be relevant to the patient population[5]. While well-designed condition-specific HRQOL instruments should provide a more sensitive response to ill-health than generic instruments, it is rare that these instruments are preference-based due to the extensive research required to generate the index value sets. Furthermore, condition-specific HRQOL instruments would not allow comparison of the value gains in HRQOL on the common metric needed by decision-making bodies such as NICE who consider health care interventions for a wide range of conditions.

In the absence of preference-based condition specific instruments, and recognising the need to measure outcomes using QALYs, techniques such as 'mapping' or 'cross-walking' can be used to predict index values from responses to condition-specific instruments[6]. A regression-based algorithm or a mapping function is used to allow index values to be predicted. The predictive quality of the algorithm depends on the conceptual overlap of the two instruments (condition-specific and preference-based). It is therefore important to assess the convergent validity of both instruments by exploring the degree to which domains are related.

The Cambridge Otology Quality of Life (COQOL) instrument is a newly developed condition-specific measure for patients with otological conditions[7], designed to be completed at routine outpatient appointments by patients who self-identify as having 'ear problems'. To place this instrument into a national framework where QALYs are routinely used to assess cost-effectiveness, this paper reports on a study that 'mapped' from COQOL to a preference-based instrument (the SF-6D). The aims of the study were: 1) to assess the convergent validity of the COQOL and the SF-6D instrument; and 2) to construct a mapping algorithm to allow QALYs to be generated from COQOL data.

METHODS

Ethical considerations

This study was devised within the University of Birmingham Health Economics Department to complement the COQOL study. Ethical approval was granted by the Cambridge Central

Ethics Committee and the Cambridge University Hospitals Research and Development Department.

Cambridge Otology Quality of Life Instrument (COQOL)

The COQOL addresses the range of symptoms experienced by patients with otological conditions. It measures quality of life with respect to ear disease and is designed to be completed by patients attending outpatient appointments. Comprising 24 items it address specific aural and auditory symptomatology (hearing in different circumstances, tinnitus, aural discharge, ear discomfort and pain) and more general quality of life fields such as self-image, inter-personal relationships and illness-behaviour. Patients are asked to rate a series of statements on a 10-point scale where 0 denotes 'strongly disagree' and 10 denotes 'strongly agree'. Each item score is added to produce an overall index score that ranges from 0 to 240. Overall, the lower the COQOL score then the better the quality of life.

Short-Form-36 Instrument (SF-36) and the SF6D

The SF-36 is a non-preference-based generic HRQOL instrument that has 36 questions covering 8 dimensions: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health[8]. The SF-6D is a preference-based health state classification system that is used to derive index values from the SF-36 instrument. Responses to six dimensions: physical functioning, role limitations (physical and emotional), bodily pain, vitality, social functioning, and mental health are used, and index values derived using a published algorithm generated from a general population sample[4]. The range of index values for the SF-6D is 0.301 to 1.00, where 1 indicates full health and 0 is death.

Both the COQOL and the SF-36 instruments were completed by 207 patients attending a routine otology clinic at Cambridge University Hospital, UK, and then again by post 3 months later. For all patients, information on age, sex, and primary pathology was collected.

Analysis

To assess the relationship between the COQOL dimensions and the SF-6D, the mean SF-6D score for different levels of response on the COQOL was examined, with the expectation

that increasing scores on each COQOL dimension (an increase in reported problems) would be associated with a lower mean SF-6D index value. For this analysis, we grouped COQOL response options into the following categories: 0-3 to reflect patients who strongly disagreed with the statement; 4-7 to reflect patients who were indifferent to the statement; and 8-10 to reflect patients that strongly agreed with the statement. Convergent validity was assessed by comparing the SF-6D mean values using a one-way ANOVA test. Furthermore, to visualise the correlation between the instruments a scatter plot of the SF-6D index values and the overall COQOL score was generated, and the correlation coefficient calculated using the Spearman's rho statistic.

We then explored the mathematical model that most accurately predicted the SF-6D index values from the COQOL questions by applying a linear regression with the SF-6D acting as dependent variable[6]. The first model focused on all questions within the COQOL, the second model replicated this, plus 'squared terms' to account for non-linearities in the variables. In both cases we used a backward selection process with statistical significance set at the 10% level[9] to select the best subset of COQOL questions to predict the SF-6D. SPSS version 20 was used for all statistical analysis.

Internal and external validation

It is good practice with any mapping exercise to test for the internal validity of the mapping models by predicting scores from within the sample, and comparing to observed values. We assessed the models on the basis of the Mean Squared Error (MSE) which is the mean of the squared differences between observed and predicted SF-6D values, and the Mean Absolute Error (MAE), the mean of absolute differences between observed and predicted SF-6D values[10]. The lower the MSE and MAE, the better the model is at predicting the index values.

To assess the external validity of the mapping models we used the data from the follow-up phase (at 3 months). Due to drop out this dataset contained 77 patients. To observe how accurately the models predict index values we ran the models and then compared the

predicted and observed index values. The same performance criteria from the internal validation exercise were then applied to this external dataset.

RESULTS

The mean age of the 207 patients in the sample at baseline was 51 years (SD 18 years) and 51% (n=105) were female. The primary pathology of the patients was stable (22%) or active (14%) chronic suppurative otitis media (CSOM), vertigo (14%), hearing loss (13%), tinnitus (8%), otsclerosis (7%), otitis externa (7%), wax (7%), Eustachian tube dysfunction (2%), or some other ontological condition (6%). An SF-6D index value could be generated for 189 (91%) of the sample, and the mean value was 0.71 (SD 0.11). The mean COQOL score (available for all patients at baseline) was 84.8 (SD 43.6). The distribution of both measures appeared approximately normal, although the SF-6D had a slight positive skew, whilst the COQOL had a slight negative skew, both toward the better quality of life end of the scale (Figure 1).

Figure 1 about here

Convergent validity

In 18/24 COQOL questions, as expected, the mean SF-6D values decreased linearly with increasing severity on the COQOL instrument, and this effect was statistically significant at the 5% level (table 1). So in these cases, the more that patients reported problems (values 8-10) using the COQOL instrument, the greater the negative impact on the mean SF-6D values. This result suggests convergent validity for these particular COQOL questions. Of interest, the negative impact on mean SF-6D values appeared to be greatest from reported problems with tinnitus, balance, pain, social life and happiness. Hearing problems seemed to affect the SF-6D values the least.

Table 1 about here

Figure 2 shows the scatter plot between the SF-6D index values and the COQOL scores. The distribution of COQOL values is much wider relative to the distribution of SF-6D as very few patients reported SF-6D values below 0.45. As expected there was a slight negative relationship with higher SF-6D values corresponding to lower COQOL scores.

Figure 2 about here

Overall, the correlation between the COQOL scores and the SF-6D values showed a moderate and statistically significant negative correlation (r = .490, p=<0.001).

The models

Table 2 shows the results for the OLS models for the prediction of the SF-6D values. According to the statistical measures (the adjusted R-squared), model 2 performed better than model 1.

Table 2 about here

Taking the results of model 1 as the simplest case, the following algorithm can be applied to COQOL data to generate index values:

SF6D index value = 0.788 + (0.010 * COQOL Q2) + (-0.008 * COQOL Q5) + (-0.011 * COQOL Q11) + (-0.008 * COQOL Q14) + (0.008 * COQOL Q15) + (-0.010 * COQOL Q21) + (-0.009 * COQOL Q24)

Internal and external validation

Table 3 displays the performance criteria for models 1 and 2 by assessing the difference between the predicted versus actual SF-6D values for the baseline sample (internal validation), and then the same again for the follow-up 3 month sample (external validation).

Table 3 about here

Both models perform well in the baseline dataset (internal test), predicting mean SF-6D values that are the same as the actual mean values. The MAE and MSE statistics also suggest that there is no difference in prediction power between the two models. In the context of the range of scores for the SF-6D (0.699; 0.301 – 1), the MAE of 0.072 to 0.079 is an average prediction error of 10-11% of the total scale of the SF-6D. Within the external dataset, both SF-6D and COQOL data were available for 77 patients at the 3-month time point. At this

follow-up the mean COQOL score was 80.8 (SD 48.1) and the SF-6D was 0.70 (SD 0.11). Again, both models performed well by predicting mean SF-6D values that are the same, or very close to, the actual SF-6D value. Again, in terms of prediction power measured by the MAE and MSE there was little difference between the models, although the MAE and MSE were closer in the internal and external validation sets for Model 1, suggesting a better model fit. The relationship between the predicted and actual SF-6D for the follow up sample is displayed in figure 3.

Figure 3 about here

DISCUSSION

Synopsis of key findings

This study has demonstrated convergent validity between the condition-specific COQOL instrument and the SF-6D preference-based instrument, and assessed how accurately SF-6D index values can be predicted from COQOL patient data. It has shown that a mapping model can predict SF-6D mean values with a low prediction error and can be used with a reasonable degree of accuracy to predict QALYs from the COQOL instrument.

Comparison with other studies

The COQOL instrument is a newly developed instrument so there are no published mapping studies to compare our results. We can however compare the predictive power of model 1 to other models that have been published predicting SF-6D mean index values for a patient population from condition specific instruments. The R^2 of 0.35 and 0.38 for models 1 and 2 respectively, is in the range of R^2 values reported for condition specific to generic instruments in a recent review of mapping studies[6]. Our results concord with what has been reported in the literature that moving to more complex model specifications such as models that include squared terms add little to the prediction power, and it is the simpler OLS models that on balance perform the best.

Strengths and weaknesses of the study

This is the first study to apply mapping techniques to a condition-specific otolaryngological quality of life instrument and provides a unique opportunity for QALYs to be generated from

COQOL otological-specific patient data. Hitherto, economic analyses of otolaryngological interventions have been limited because of the need for data that is required to measure QALYs [11].

One drawback of the study was the sample as although typical of patients who present at general otology clinics, the numbers were relatively small and ideally a separate patient sample should have been used for the external validation. Instead the best data available was used and future research can replicate this analysis on a separate patient sample to further assess the prediction power of the mapping algorithm. COQOL and SF-6D values also indicated that the sample used in this study had relatively high HRQOL so ideally the performance of the algorithm needs further validation for use in populations with greater HRQOL impairments.

Clinical applicability of the study

For decision makers interested in the cost-effectiveness of otological treatments, this study presents a mapping algorithm to generate mean SF-6D index values from COQOL data and is therefore useful where no generic preference-based data exists but QALY outcomes are beneficial. However, although mapping offers a means to generate QALYs, it should always be viewed as a second-best option to collecting preference-based measures, such as the SF-6D, which allow index values to be generated directly from patient responses.

Reference List

(1) National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2008.

(2) Neumann P, Greenberg D. Is the United States ready for QALYs? *Health Affairs* 2009; 28(5):1366-1371.

(3) Brooks R, Rabin R, de Charro F. The measurement and valuation of health status using EQ-5D: a European perspective. Kluwer Academic Publishers; 2003.

(4) Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002; 21(2):271-292.

(5) Fitzpatrick R, Davey C, Buxton MJ, Jones D. Evaluating patient-based outcome measures for use in clinical trials. 2(14). 1998. Health Technology Assessment.

(6) Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) from non-preference based measures of health to generic preference based measures. *European Journal of Health Economics* 2010; 11:215-225.

(7) Martin TPC, Mouled D, Paul A, Ronan N, Tysome J, Donnelly N et al. The Cambridge Otology Quality of Life Questionnaire: an otology-specific patient recorded outcome measure. A paper describing the instrument design and a report of preliminary reliability and validity. *Clinical Otolaryngology*, in press 2014.

(8) Ware JE, Snow KK, Kosinski M, Gandek B. SF36 Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Centre, The Health Institute; 1993.

(9) Collet D. Modelling Survival Data in Medical Research. New York: Chapman and Hall/CRC; 1994.

(10) Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value in Health* 2013; 16(1):202-210.

(11) Gait C, Frew EJ, Martin TPC, Jowett S, Irving R. Conservative management, surgery, and radiosurgery for treatment of vestibular schwannomas: cost effectiveness analysis. *Clinical Otolaryngology* 2014; 39(1):22-31.

Table 1: Mean SF-6D index value by each level of COQOL question

COQOL Questions	Level	N	Mean SF6D score (SD)	P*
1) I struggle to hear what people are saying to me.	0 - 3 4 – 7 8 – 10	51 84 52	0.72 (0.12) 0.71 (0.12) 0.68 (0.11)	.353
2) If I'm in a group of more than 3 people, I find it hard to follow the conversation.	0 - 3 4 – 7 8 – 10	55 60 71	0.71 (0.12) 0.69 (0.12) 0.71 (0.11)	.647
3) I find it hard to tell if a car is coming towards me or going away from me.	0 - 3 4 – 7 8 – 10	93 56 40	0.71 (0.11) 0.69 (0.12) 0.69 (0.12)	.561
4) Music doesn't sound clear or natural to me.	0 - 3 4 – 7 8 - 10	86 74 29	0.72 (0.11) 0.69 (0.11) 0.67 (0.12)	.066
5) People around me have told me that I turn the TV or radio up too loud.	0 - 3 4 – 7 8 - 10	50 47 88	0.71 (0.12) 0.73 (0.11) 0.68 (0.11)	.048
6) I struggle to use the telephone	0 - 3 4 – 7 8 – 10	88 61 35	0.73 (0.11) 0.67 (0.12) 0.69 (0.11)	.032
7) Tinnitus keeps me awake at night	0 - 3 4 – 7 8 – 10	117 36 30	0.72 (0.10) 0.69 (0.13) 0.64 (0.12)	.004
8) My tinnitus makes it difficult for me to follow a conversation in background noise.	0 - 3 4 – 7 8 – 10	113 37 30	0.72 (0.10) 0.68 (0.13) 0.66 (0.12)	.023
9) I am bothered by my tinnitus during the day-time.	0 - 3 4 – 7 8 – 10	114 35 30	0.73 (0.11) 0.67 (0.11) 0.65 (0.13)	.002
10) Sometimes I feel so dizzy that I can't do anything at all.	0 - 3 4 – 7 8 – 10	128 24 31	0.73 (0.10) 0.64 (0.12) 0.65 (0.13)	.000
11) My daily activities are limited by my balance.	0 - 3 4 – 7 8 – 10	138 31 15	0.73 (0.10) 0.64 (0.12) 0.58 (0.12)	.000
12) I am troubled by clicking and popping sounds in my ears.	0 - 3 4 – 7 8 – 10	120 35 27	0.72 (0.10) 0.68 (0.13) 0.65 (0.12)	.010
13) My ears (or ear) feel blocked.	0 - 3 4 – 7 8 – 10	71 49 65	0.73 (0.11) 0.72 (0.12) 0.67 (0.11)	.002
14) My ears (or ear) are painful.	0 - 3 4 – 7 8 – 10	111 52 22	0.73 (0.10) 0.69 (0.11) 0.61 (0.14)	.000
15) I suffer from smelly discharge from my ears.	0 - 3 4 – 7 8 – 10	142 18 25	0.71 (0.11) 0.72 (0.13) 0.68 (0.12)	.427
16) I am embarrassed by my ear discharge.	0 - 3 4 – 7 8 – 10	142 16 20	0.72 (0.10) 0.68 (0.15) 0.69 (0.11)	.254
17) I often need to visit the doctor about my ears.	0 - 3 4 – 7 8 – 10	96 44 42	0.73 (0.10) 0.67 (0.14) 0.69 (0.10)	.015
18) I often feel unwell because of my ears.	0 - 3	114	0.74 (0.10)	.000

	4 – 7 8 – 10	40 32	0.68 (0.12) 0.61 (0.12)	
19) My work or other responsibilities are affected by my problems with my ears.	0 - 3 4 – 7 8 – 10	92 38 55	0.74 (0.10) 0.72 (0.10) 0.64 (0.11)	.000
20) I feel that my ear problems affect my concentration.	0 - 3 4 – 7 8 – 10	95 46 45	0.75 (0.10) 0.67 (0.11) 0.64 (0.11)	.000
21) I feel limited in my social life because of my ear problems.	0 - 3 4 – 7 8 – 10	104 40 43	0.74 (0.10) 0.67 (0.12) 0.64 (0.11)	.000
22) I often feel unhappy because of my ear problems.	0 - 3 4 – 7 8 – 10	85 52 49	0.75 (0.10) 0.68 (0.11) 0.66 (0.11)	.000
23) My ear problems affect how I spend my leisure time.	0 - 3 4 – 7 8 – 10	97 42 46	0.75 (0.10) 0.67 (0.12) 0.63 (0.10)	.000
24) My ear problems have affected my sense of taste.	0 - 3 4 – 7 8 – 10	134 31 17	0.73 (0.10) 0.63 (0.13) 0.62 (0.10)	.000

Table 2: Mathematical models for predicting SF6D index values

	OLS	
	(1)	(2)
COQOL questions:		
I find it hard to follow the conversation (Q2)	0.010	0.011
I turn the TV or radio up too loud (Q5)	-0.008	-
My daily activities are limited by my balance (Q11)	-0.011	-0.010
My ears (or ear) are painful (Q14)	-0.008	-
I suffer from smelly discharge from my ears (Q15)	0.008	-
I feel limited in my social life (Q21)	-0.010	-
My ear problems have affected my sense of taste (Q24)	-0.009	-0.010
I often need to visit the doctor about my ears (Q17)	-	-0.017
COQOL questions squared:		
I turn the TV or radio up too loud (Q5)	-	-0.001
My ears (or ear) are painful (Q14)	-	-0.001
I suffer from smelly discharge from my ears (Q15)	-	0.001
I often need to visit the doctor about my ears (Q17)	-	0.002
I feel limited in my social life (Q21)	-	-0.001
Constant	0.788	0.777
R-squared	0.383	0.419
Adjusted R-squared	0.356	0.383

Table 3: Performance of models

	Internal validation		External validation	
	Model 1	Model 2	Model 1	Model 2
SF-6D: Actual, mean (SD)	0.71 (0.12)	0.71 (0.12)	0.70 (0.11)	0.70 (0.11)
SF-6D: Predicted, mean (SD)	0.71 (0.07)	0.71 (0.08)	0.71 (0.08)	0.70 (0.09)
MAE*	0.072	0.072	0.073	0.079
MSE*	0.008	0.008	0.008	0.010
Predictions within 0.1 and 0.05 of observed values:				
0.05	44%	44%	50%	43%
0.1	78%	77%	76%	71%

Figure 1: Histogram of COQOL and SF-6D data for the baseline sample

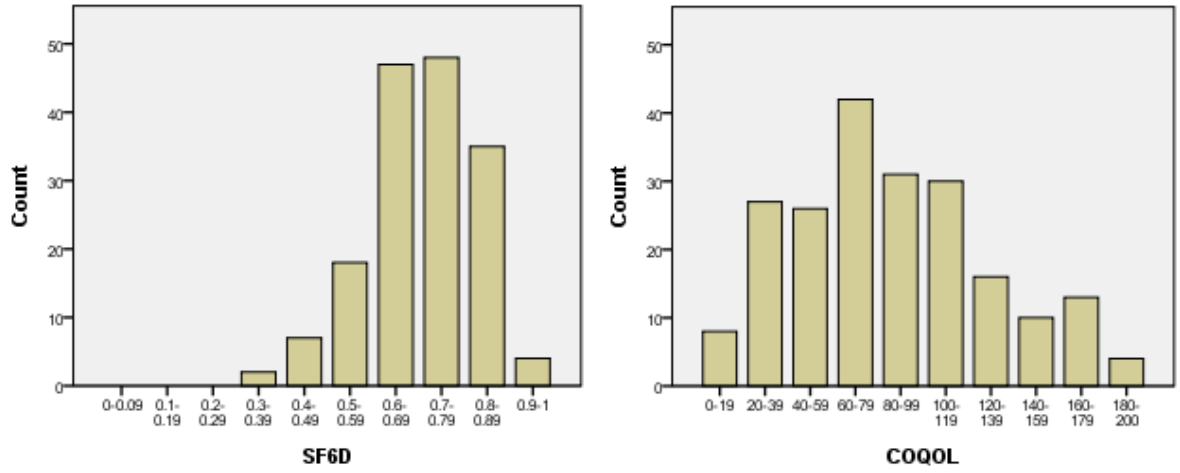

Figure 2 Scatter plot of COQOL scores and SF-6D values at baseline (first wave)

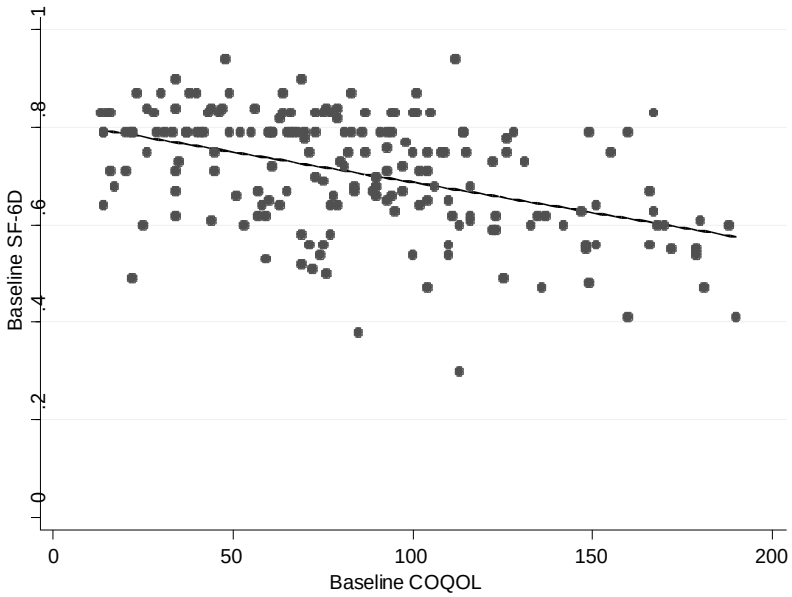

Figure 3: Follow up observed and predicted SF6D values for model 1 and model 2